

A close-up photograph of two women's faces and hands. The woman on the left has blonde hair and is wearing a yellow patterned top, smiling while holding a gold smartphone. The woman on the right has dark hair and is wearing a blue and white patterned top, also smiling and holding a red smartphone. A thin black horizontal line runs across the middle of the slide.

DRESS LIKE A STAR: RETRIEVING FASHION PRODUCTS FROM VIDEOS

NOA GARCIA & GEORGE VOGIATZIS

A large, blurry background image showing the same two women from the previous slide, now seen from the waist up. They are both looking down at their smartphones. The woman on the left is wearing a ring on her left hand, and the woman on the right has pink nail polish.

COMPUTER VISION IN FASHION WORKSHOP

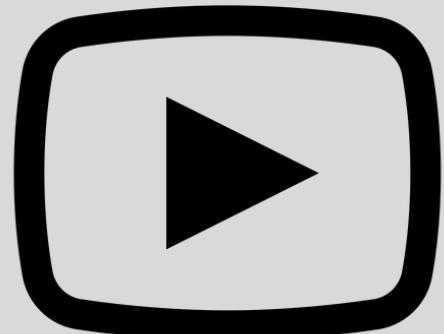
Fashion in Videos



Movies



TV shows



Online

Fashion in Videos



Sex and the City

Fashion in Videos



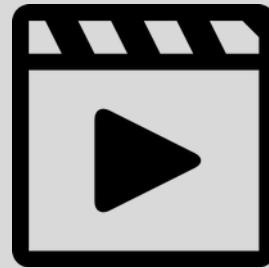
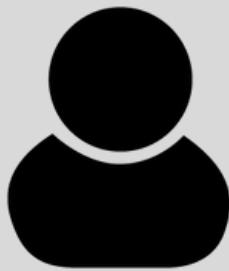
The Devil Wears Prada

Fashion in Videos



The Great Gastby

Fashion in Videos



Make fashion products in videos more accessible to users.

Fashion in Videos



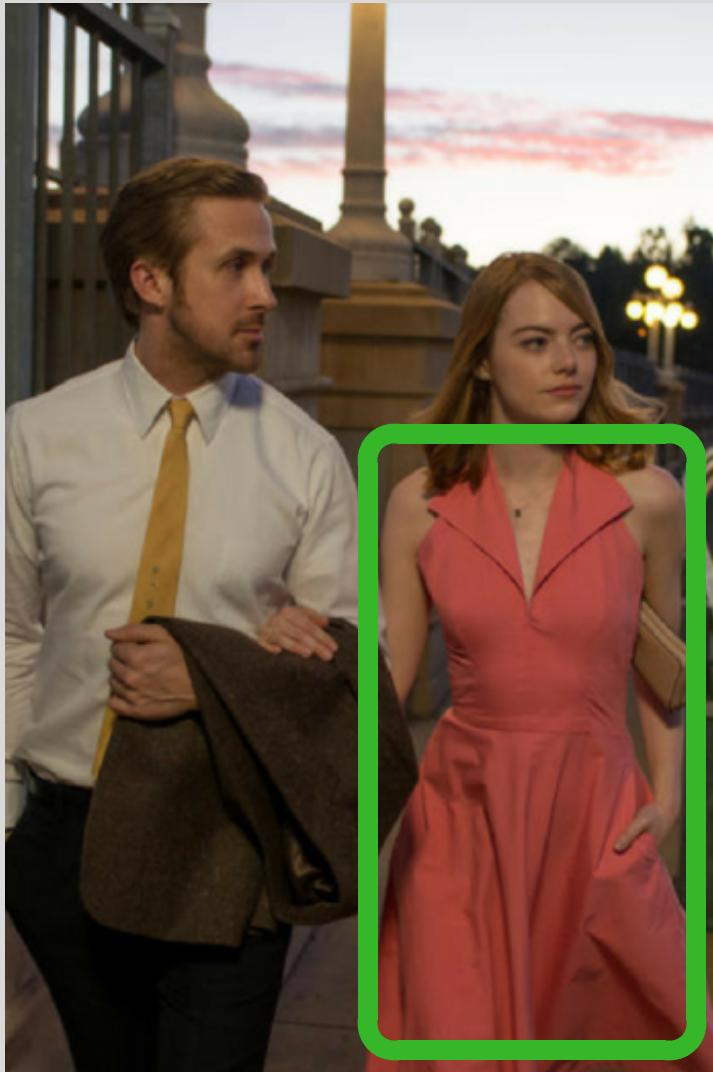
Constraints

1. Camera view



Camera viewpoint cannot be moved to have a better view of the fashion object.

Constraints

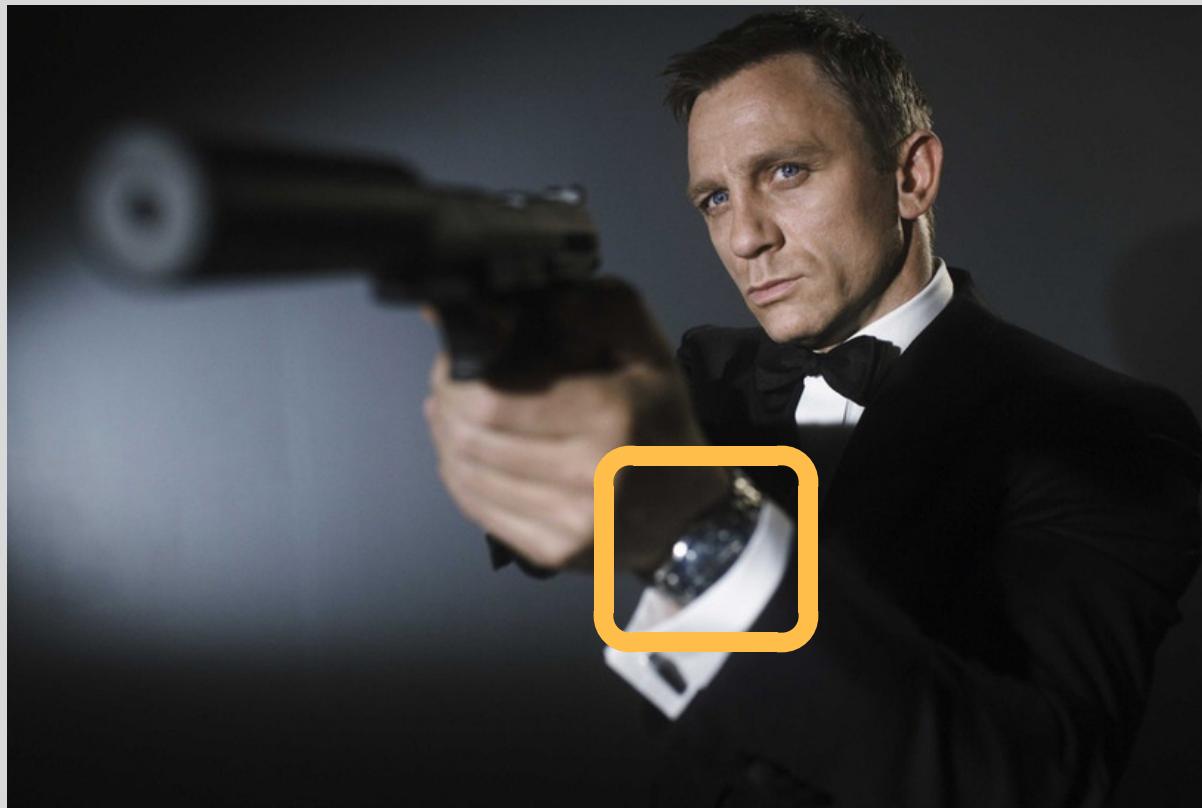


2. User interaction

The creation of bounding boxes around the object of interest may distract users from the video.

Constraints

3. Small objects



Small,
partially
occluded and
blurred.

Our Proposal

Instead of object recognition...



Our Proposal

Instead of object recognition... frame retrieval



Related Work

Clothing Retrieval

- Attribute classification [1]
- Domain adaptation [2]

Scene Retrieval

- Image Retrieval in Videos [3]
- Temporal tracking [4]
- Scene Descriptors [5, 6]

Our Approach: binary temporal tracking + fast indexing.

Challenges

- Average movie duration **120 minutes**
- Standard FPS rate **24 fps**
- Average frames per movie **172,800 frames**

With only 5 or 6 movies



More than a million frames!

Our System

Three main modules:

- Product indexing
- Training phase
- Query phase

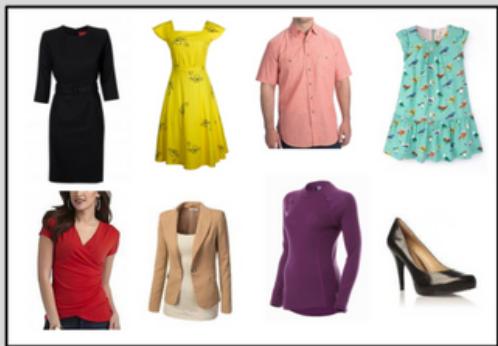
Our System

Three main modules:

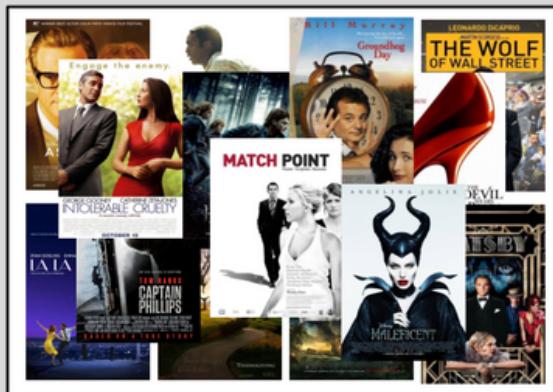
- Product indexing
- Training phase
- Query phase

Our System: Product indexing

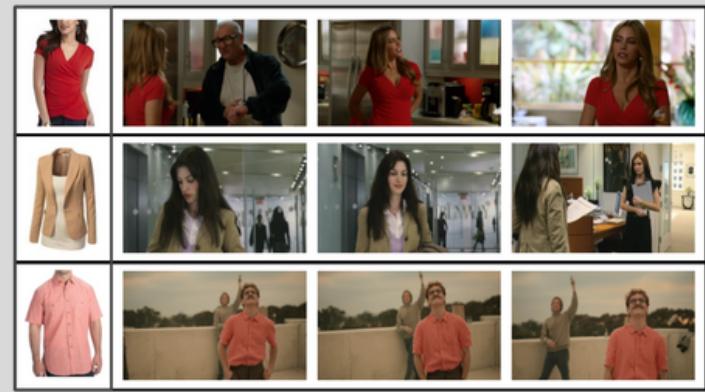
Product Indexing



Fashion Items



Video Collection



Indexed Database

Fashion items and frames related in an database.

Our System

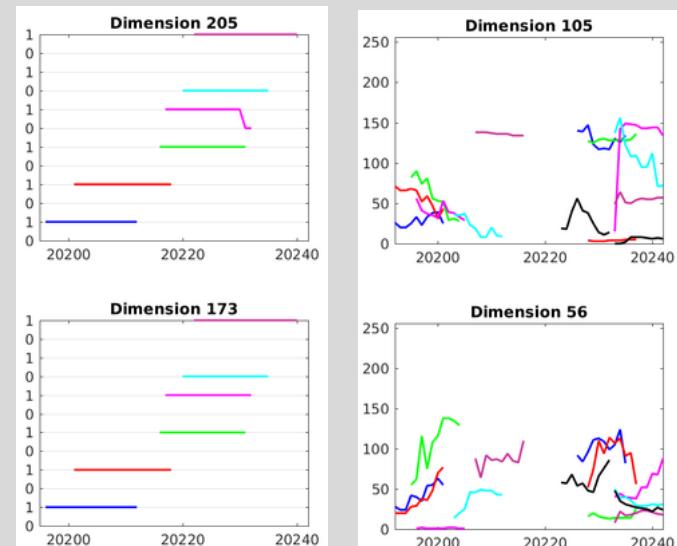
Three main modules:

- Product indexing
- Training phase
- Query phase

Our System: Training phase



BRIEF features are more constant over time than **SIFT** or **CNN**.



BRIEF

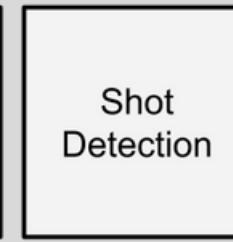
SIFT

Our System: Training phase

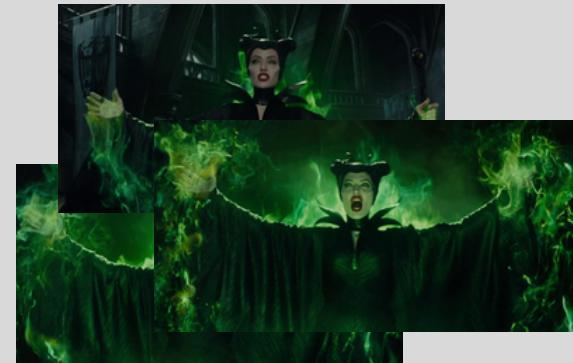
Training phase



Video Frames



shot 1



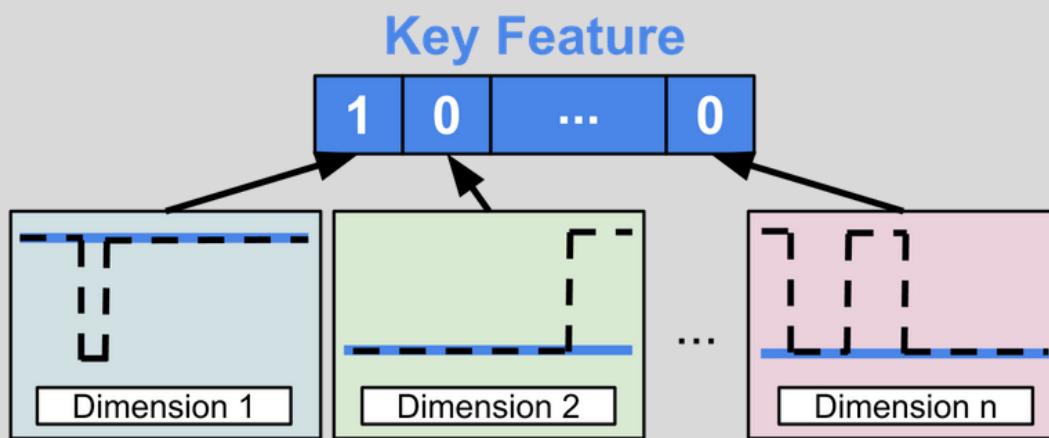
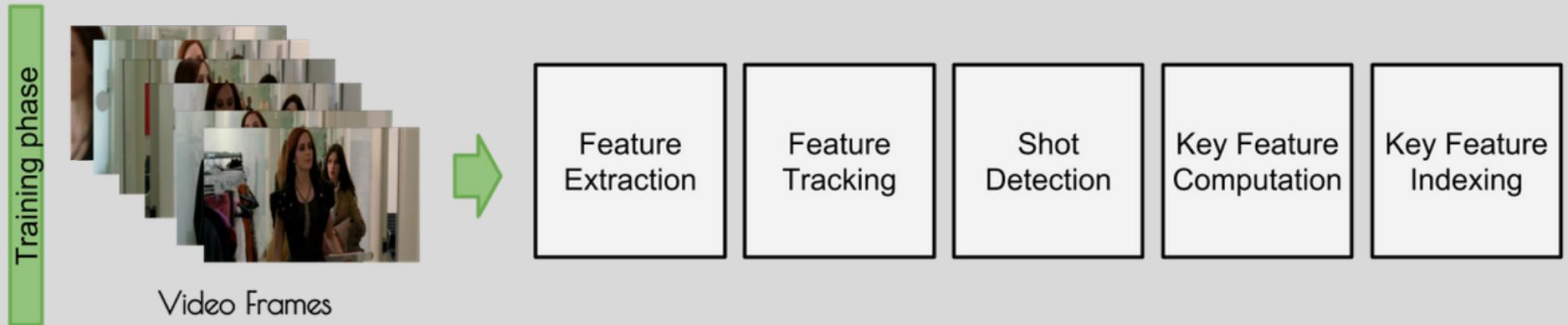
shot 2



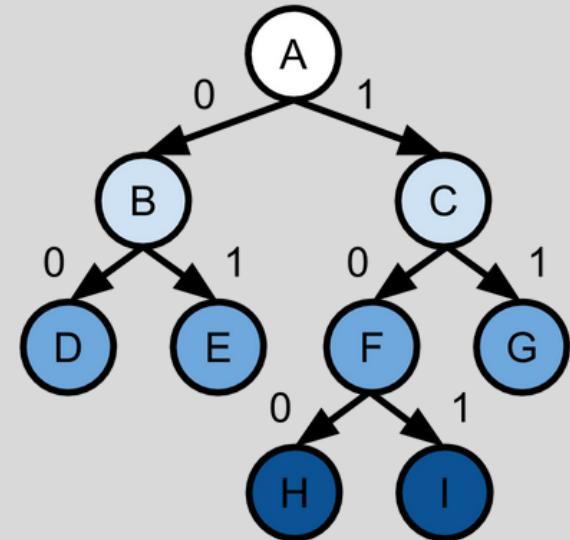
shot 3

Similar frames are grouped into shots.

Our System: Training phase



Kd-Tree Indexing

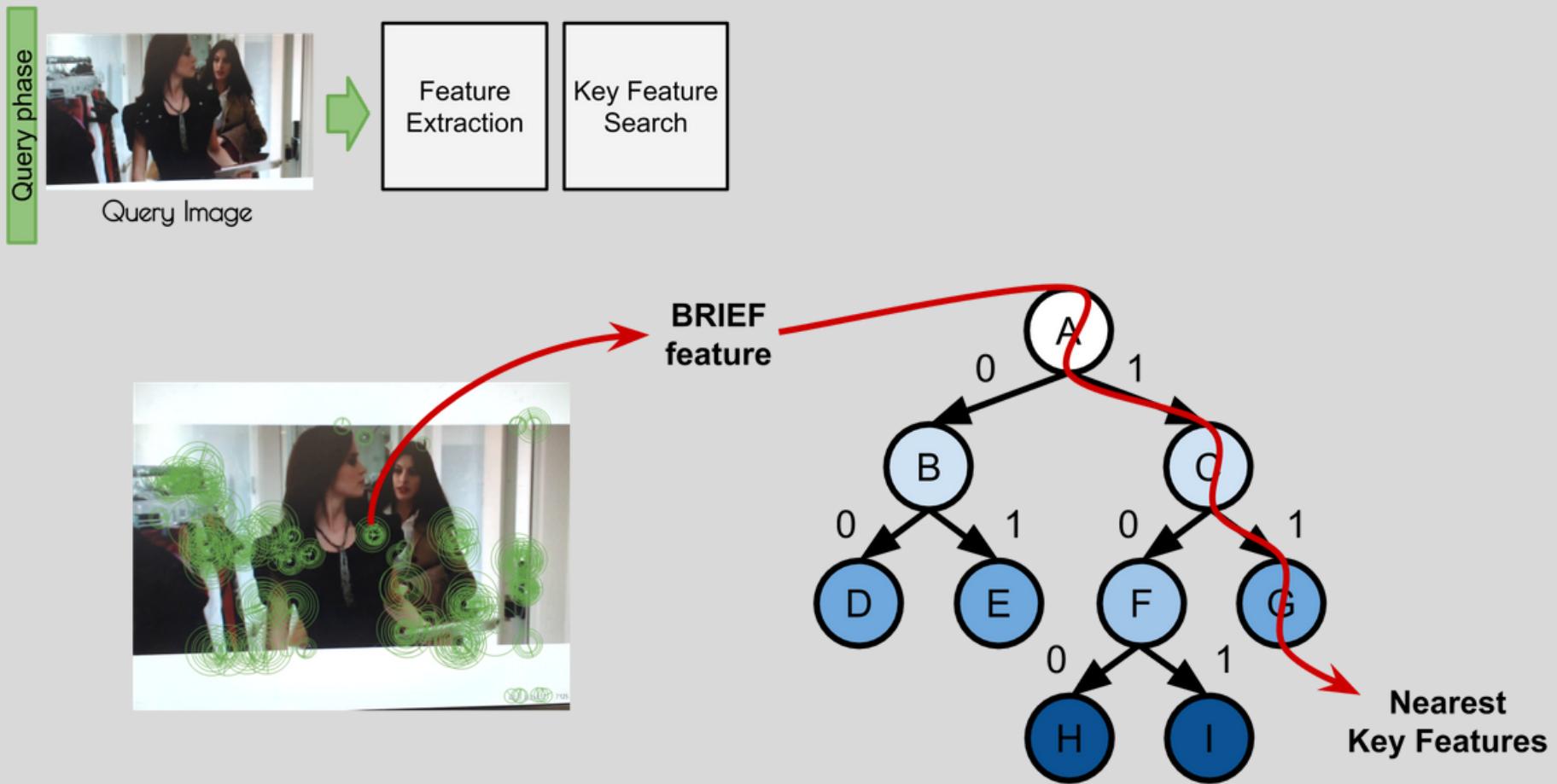


Our System

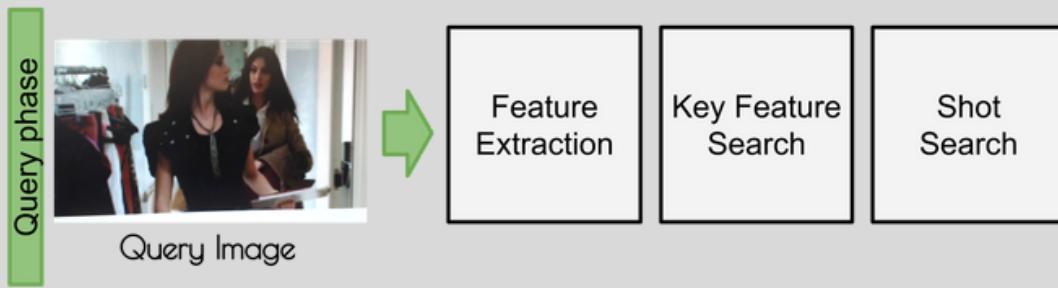
Three main modules:

- Product indexing
- Training phase
- Query phase

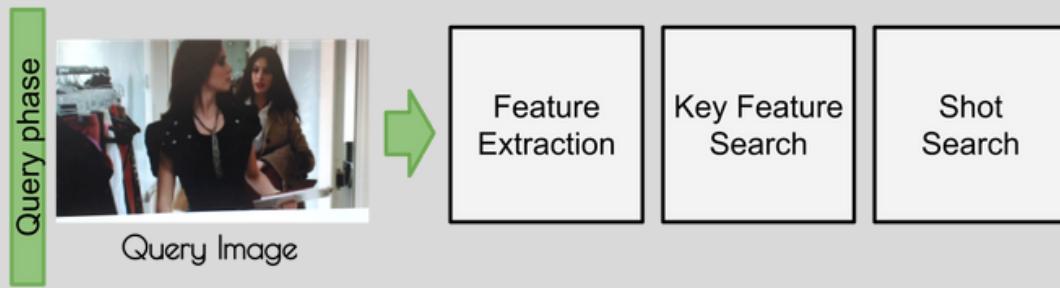
Our System: Query phase



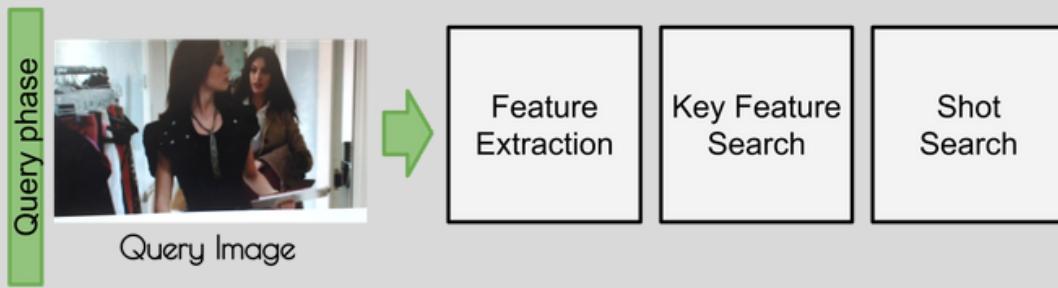
Our System: Query phase



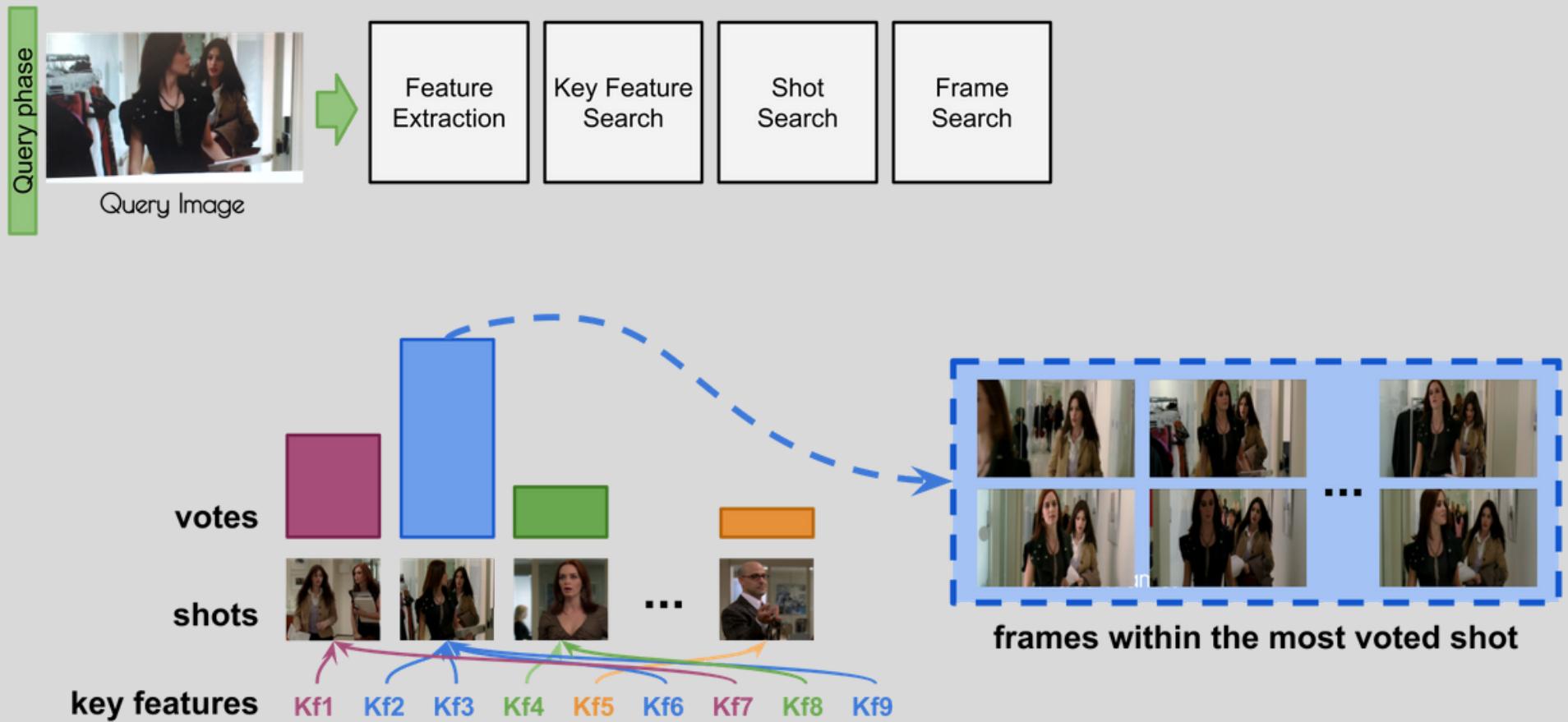
Our System: Query phase



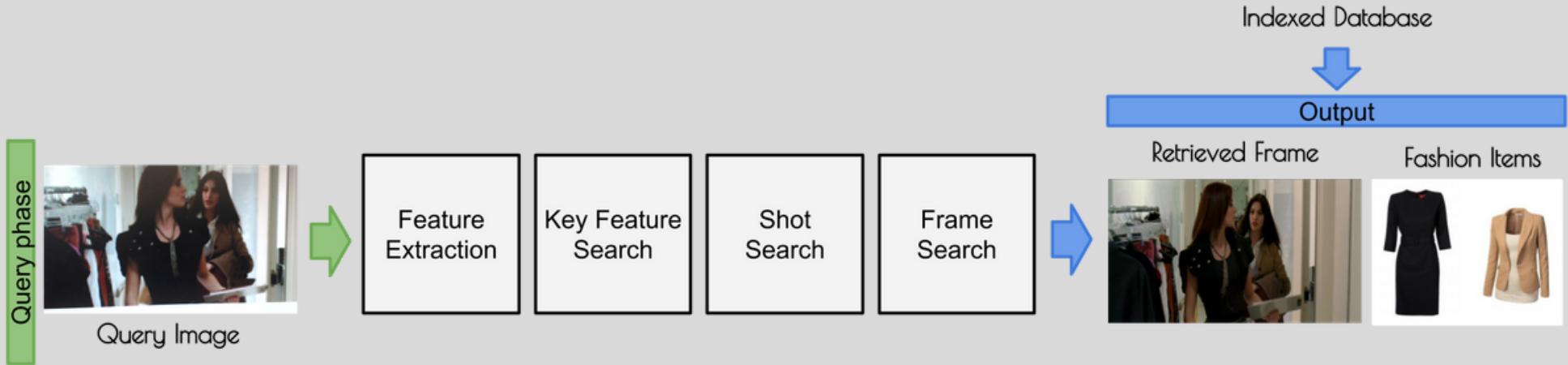
Our System: Query phase



Our System: Query phase



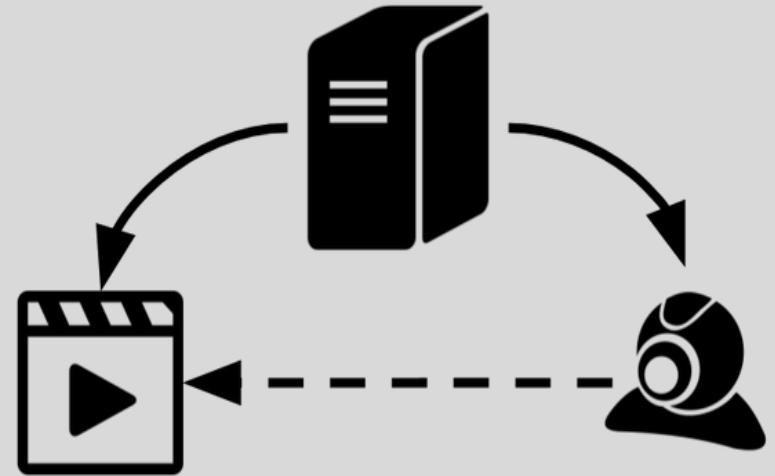
Our System: Query phase



Use the most similar frame to find the fashion products in the indexed product database.

Experiments - Dataset

- Webcam captures video playback.
- Frame number is used as a ground truth.



The retrieved frame should be visually similar to the annotated ground truth.

Experiments - Retrieval Performance

Results using a single movie, 1h 49min duration

Huge gain in
memory
requirements
with our method.

BF: Brute Force
KT: Kd-Tree
KF: Key Frame

		BF	KT	KF	Ours
Indexed Features		85M	85M	25M	2M
Memory		2.53GB	2.53GB	762MB	61MB
Accuracy	B = 10	0.98	0.90	0.91	0.92
	B = 50		0.94	0.92	0.93
	B = 100		0.96	0.93	0.94
	B = 250		0.97	0.93	0.94

Experiments - Scalability

40 movies
80 hours
7 million frames

A word cloud visualization where movie titles and subtitles are represented by colored text. The size of the text indicates its frequency or importance. The titles include: The Social Network, The Wolf of Wall Street, Absolutely Anything, Grave of the Fireflies, Pirates of the Caribbean, The Help, American Hustle, Captain Phillips, Spanish Affair 2, Marshland, Lee Daniels' The Family United, Magnolia, Her, El Niño, Casablanca, 300: Rise of an Empire, Witching and Bitching, Neon Genesis Evangelion, The Great Gatsby, 2 Francs, 40 Pesetas, Match Point, Despicable Me, Puss in Boots, A Single Man, The Physician, Maleficent, Seven Pounds, Rise of the Planet of the Apes, Out of Africa, The Hobbit: The Desolation of Smaug, Big Fish, Groundhog Day, The Body, 12 Years a Slave, Ant-Man, Harry Potter and the Deathly Hallows, The Devil Wears Prada.

Experiments - Scalability

Results using 40 movies

Title	N. Frames	N. Features	N. Shots	N. Key Features	N. Queries	Accuracy
The Help	210387	101M	1726	2.2M	813	0.98
Intolerable Cruelty	179234	86M	1306	2M	544	0.97
Casablanca	147483	71M	881	1.5M	565	0.96
Witching & Bitching	163069	66M	4193	0.8M	588	0.74
Pirates of the Caribbean 3	241127	108M	3695	1.7M	881	0.74
Captain Phillips	190496	59M	7578	0.6M	618	0.67
Total	7M	3040M	116307	58M	25142	0.87

Data reduction:

From 3,040M features to 58M key features.

Conclusions

- System to perform video clothing retrieval.
- It helps users to find items shown in videos.
- Based on frame retrieval and fast indexing.
- It scales well when the collection is increased.

THANK YOU!

NOA GARCIA
ASTON UNIVERSITY

CONTACT: GARCIA.DN@ASTON.AC.UK
GITHUB: NOAGARCIA/DRESSTAR

COMPUTER VISION IN FASHION WORKSHOP

References

- [1] Z. Liu, P. Luo, S. Qiu, X. Wang, and X. Tang. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In CVPR, 2016.
- [2] S. Liu, Z. Song, G. Liu, C. Xu, H. Lu, and S. Yan. Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set. In CVPR, 2012.
- [3] J. Sivic and A. Zisserman. Video Google: a text retrieval approach to object matching in videos. In ICCV, 2003.
- [4] A. Anjulan and N. Canagarajah. Object based video retrieval with local region tracking. Signal Processing: Image Communication, 22(7), 2007.
- [5] C.-Z. Zhu and S. Satoh. Large vocabulary quantization for searching instances from videos. In ACM ICMR, 2012.
- [6] A. Araujo and B. Girod. Large-scale video retrieval using image queries. IEEE Transactions on Circuits and Systems for Video Technology, 2017.